

# Metadata Evaluation and Improvement: Evolving Analysis and Reporting

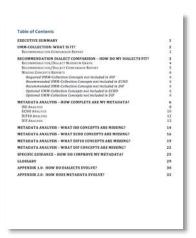
Ted Habermann, John Kozimor, Sean Gordon The HDF Group thabermann@hdfgroup.org



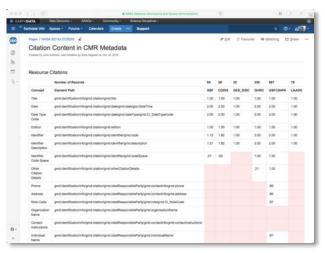
#### Metadata Analysis / Reporting Evolution



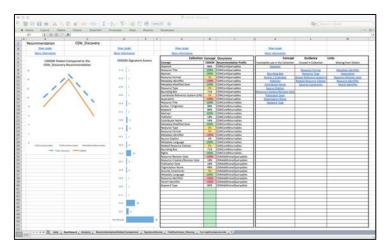
The collections we are analyzing are becoming more detailed (CMR -> DAAC) and our tools are becoming more detailed and interactive



Report (doc)



Quick Evaluation (wiki)



Dashboard (excel)

#### Terminology

**Collection:** A group of metadata records, commonly organized by a data center, organization or project and often stored in a database or web accessible folder.

**Concept:** General term for describing a documentation entity (e.g. Title, Revision Date, Process Step, Spatial Extent).

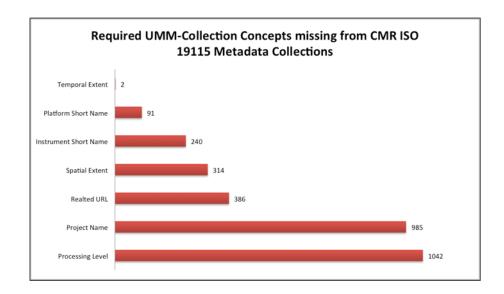
**Dialect:** A particular form of the documentation language that is specific to a community (e.g. DIF, CSDGM, EML, ECHO).

**Recommendation:** A set of concepts that a group believes is required for achieving a documentation goal.

**Spiral:** A set of concepts required to support a particular documentation need or use case for a recommendation.

#### Evaluation Report (i.e. Case-Study)

- Comprehensive completeness evaluation for multiple dialects and recommendations.
- Analyze complete metadata collections of any size. (no processing limitations).
- **Deliverable is a comprehensive report** which includes detailed analysis, supported by a variety of quantitative visualizations and a narrative interpretation of the results.
- Often includes an appendix with more detailed exploratory analysis.
- This report may take weeks to compile and draft. More comprehensive, but less agile then the RAD report.
- ESDIS reports covered the entire CMR collection no differentiation by DAAC.



CONCEPT	DIF	DIF10	ECHO	ISO19115
Purpose	0%	7%	88%	88%
Resource Language	100%	91%	*	0%
Resource Contact <sub>3</sub>	84%	84%	0%	9%
Responsibility	0%	0%	0%	0%
Collection Data Type <sub>4</sub>	95%	95%	95%	100%
Resource Citation <sub>5</sub>	0%	91%	0%	9%
Resource Status <sub>6</sub>	39%	39%	33%	42%
Quality Statement <sub>7</sub>	0%	5%	82%	84%
Resource Use Constraints <sub>8</sub>	0%	5%	62%	0%
Resource Access Constraints	100%	95%	0%	0%
Topic Category	100%	91%	*	100%
Media	100%	91%	*	100%
Resource Format	100%	91%	61%	61%
Transfer Size	100%	99%	*	100%
Resource Cost or Fees	100%	91%	48%	48%

#### **Quick Evaluation Report**

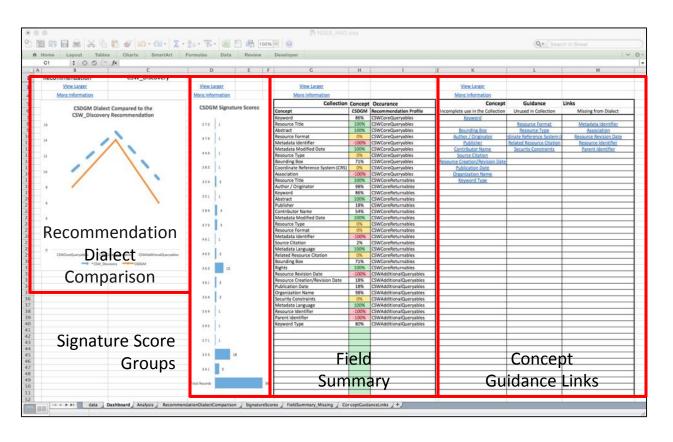
- Empirical analysis of fields (XML) used in metadata collections (no recommendations).
- Comparisons across collections (i.e. DAACs).
- Identifies fields organizations consider most important (bottom up empirical approach).
- Detailed view of the metadata. All fields in an object are compared.
- See https://wiki.earthdata.nasa.gov/display/NASAISO/NASA+ISO+for+EOSDIS

#### Occurrences of Metadata Contact Responsibility

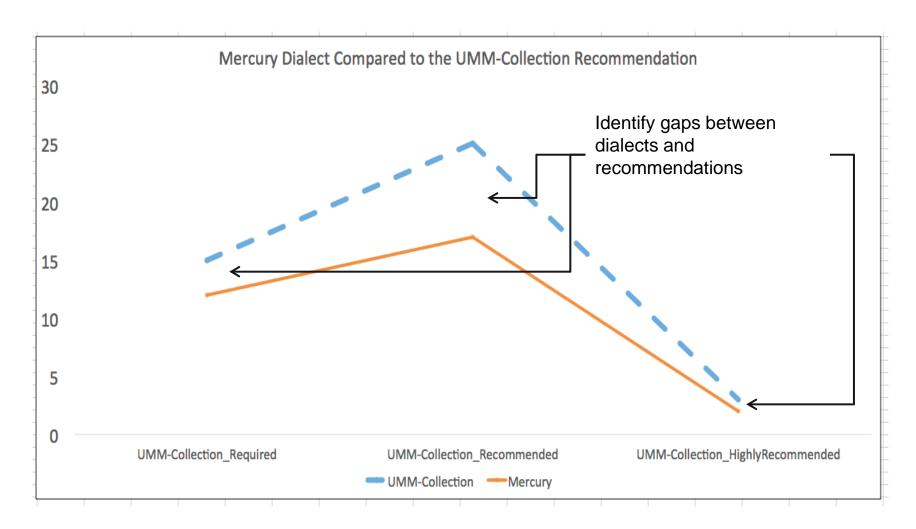
	Number of Records		84	28	35	5	339	9 68	87	78	5		11	6	343	495	18	15	197	63		4	1135	37	5 11	
Concept gmd:organisationNa	Platform Content in Created by John Kazimor, last modified by Re																'									
	Number of Records									32	36	267	491	53	43	319	134	164	280	77	68	973	80	170	11	2158
gmd:CI_RoleCode	Concept						ASF	CDDIS	GHRC	GSFCS4PA	LAADS	LANCEMODIS	LARC	LARC_ASDC		NSIDCV0	NSIDC ECS	OBPG		_		USGS_EROS				
gmd:individualName	Platform ID						1.00	7.14	1.27	.85	1.00	1.00	1.00	6.30	1.13	1.62	1.47	1.00	1.39	1.93	SEDAC	2.00	# Collectio			
gmd:deliveryPoint							_													_		2.00				
,	Platform Identifier Description					1.00	7.14	1.27	.85	1.00	1.00	1.00	6.30	1.13	1.62	1.47	1.00	1.39	1.93			16				
gmd:city	Platform Description	gmi:acquisitionInfor			-					1.00	.06	1.27	.85	1.00	1.00	1.00	6.30	1.13	1.62	1.47	1.00	1.39	1.93		2.00	15
gilla.city	Platform Identifier					1.00	7.14	1.27	.85	1.00	1.00	1.00	6.30	1.13	1.62	1.47	1.00	1.39	1.93		2.00	15				
	Platform Additional Attribute Name	gmi:acquisitionInformation/gmi:platform/eos:otherProperty/gco:Record /eos:AdditionalAttributes/eos:AdditionalAttribute/eos:reference/eos:name								.24	.98	1.00	.75		1.07		1.99					2.00	7			
gmd:administrativeA	Platform Additional Attribute Code List	gml:acquisitionInformation/gml:platform/eos:otherProperty/goo:Record /eos:AdditionalAttributes/eos:AdditionalAttribute/eos:reference/eos:type/@codeList							.24	.98	1.00	.75		1.07		1.99					2.00	7				
gmd:postalCode	Platform Additional Attribute Code List Value	gmi:acquisitionInfon /eos:AdditionalAttrib							ListValue				.24	.98	1.00	.75		1.07		1.99					2.00	7
gmd:country	Platform Additional Attribute Type Code	gmi:acquisitionInformation(gmi:platform/eos:otherProperty/gco:Record /eos:AdditionalAttributes/eos:AdditionalAttribute/eos:reference/eos:type /eos:EOS_AdditionalAttributeTypeCode							.24	.98	1.00	.75		1.07		1.99					2.00	7				
gmd:voice	Platform Additional Attribute Value	gmi:acquisitionInfon /eos:AdditionalAttrib					gco:Reco	rd					.24	.98	1.00	.75		1.07		1.99					2.00	7
	Platform Other Record Type	gmi:acquisitionInfor	mation/gr	mi:platfo	rm/eos:oth	erPropertyT	ype/gco:F	RecordType	е				.05	.98	1.00	.33		1.07		.57					2.00	7
	Platform Other Record Reference	gmi:acquisitionInfor	mation/gr	mi:platfo	rm/eos:oth	erPropertyT	ype/gco:F	RecordType	e/@xlink:hre				.05	.98	1.00	.33		1.07		.57					2.00	7
	Platform Additional Attribute Data Type Code List	gmi:acquisitionInfon /eos:AdditionalAttrib							codeList					.98	1.00	.45		1.07		.88					2.00	6
	Platform Additional Attribute Data Type Code List Value	gmi:acquisitionInfon /eos:AdditionalAttrib /@codeListValue												.98	1.00	.45		1.07		.88					2.00	6
	Platform Additional Attribute Data Type Code	gmi:acquisitionInfon /eos:AdditionalAttrib /eos:EOS_Additional	outes/eos	Additio	nalAttribute									.98	1.00	.45		1.07		.88					2.00	6
	Platform Additional Attribute Description	gmi:acquisitionInfon /eos:AdditionalAttrib												.98	1.00	.45		1.07		.88					2.00	6
	PlatformAdditionalAttribute Measurement Units	gmi:acquisitionInfon /eos:AdditionalAttrib /eos:parameterUnits	outes/eos	Additio				rd						.98	1.00	.45		1.07		.88					2.00	6

#### Recommendations Analysis Dashboard

- Interactive exploratory **metadata concept** evaluation tool.
- Enables metadata for a single dialect to be easily evaluated using multiple recommendations (eg. CSW, DataCite).
- Designed to run on collections of ~100 metadata records (samples of collections).
- Provides a nice dashboard interface with 4 different displays.
- Requires a data sheet, created by HDF metadata team.



#### Recommendation / Dialect Comparison



#### Field Summary

Identify fields that are missing from dialect, missing from collection, complete, or partial

-100% = Concept Not in Dialect

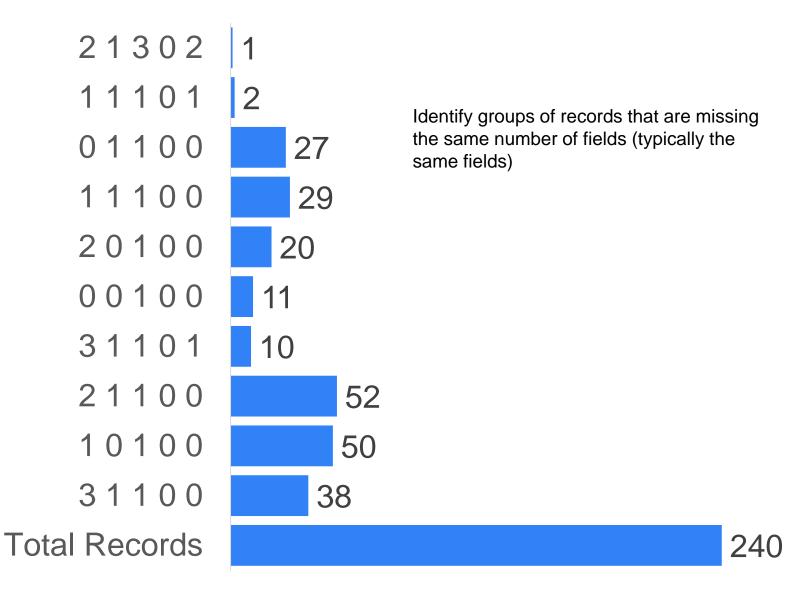
0% = Concept Not in Collection

100% = Concept in All Records

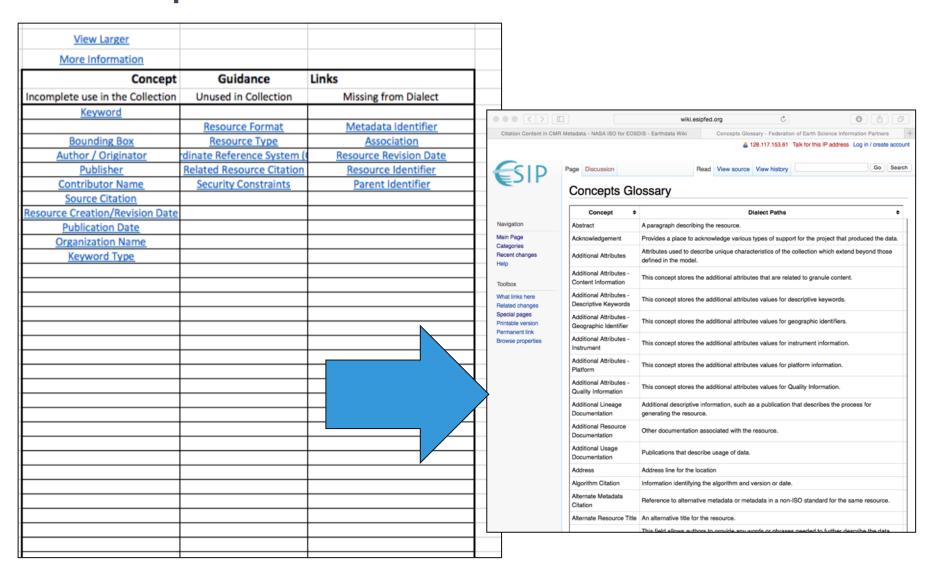
54% = Concept in Some Records

View Larger		
More Information		
Collection	Concept	Occurance
Concept	CSDGM	Recommendation Profile
Keyword	86%	CSWCoreQueryables
Resource Title	100%	CSWCoreQueryables
Abstract	100%	CSWCoreQueryables
Resource Format	0%	CSWCoreQueryables
Metadata Identifier	-100%	CSWCoreQueryables
Metadata Modified Date	100%	CSWCoreQueryables
Resource Type	0%	CSWCoreQueryables
Bounding Box	71%	CSWCoreQueryables
Coordinate Reference System (CRS)	0%	CSWCoreQueryables
Association	-100%	CSWCoreQueryables
Resource Title	100%	CSWCoreReturnables
Author / Originator	98%	CSWCoreReturnables
Keyword	86%	CSWCoreReturnables
Abstract	100%	CSWCoreReturnables
Publisher	18%	CSWCoreReturnables
Contributor Name	54%	CSWCoreReturnables
Metadata Modified Date	100%	CSWCoreReturnables
Resource Type	0%	CSWCoreReturnables
Resource Format	0%	CSWCoreReturnables
Metadata Identifier	-100%	CSWCoreReturnables
Source Citation	2%	CSWCoreReturnables
Metadata Language	100%	CSWCoreReturnables
Related Resource Citation	0%	CSWCoreReturnables
Bounding Box	71%	CSWCoreReturnables
Rights	100%	CSWCoreReturnables
Resource Revision Date	-100%	CSWAdditionalQueryables
Resource Creation/Revision Date	18%	CSWAdditionalQueryables
Publication Date	18%	CSWAdditionalQueryables
Organization Name	98%	CSWAdditionalQueryables
Security Constraints	0%	CSWAdditionalQueryables
Metadata Language	100%	CSWAdditionalQueryables

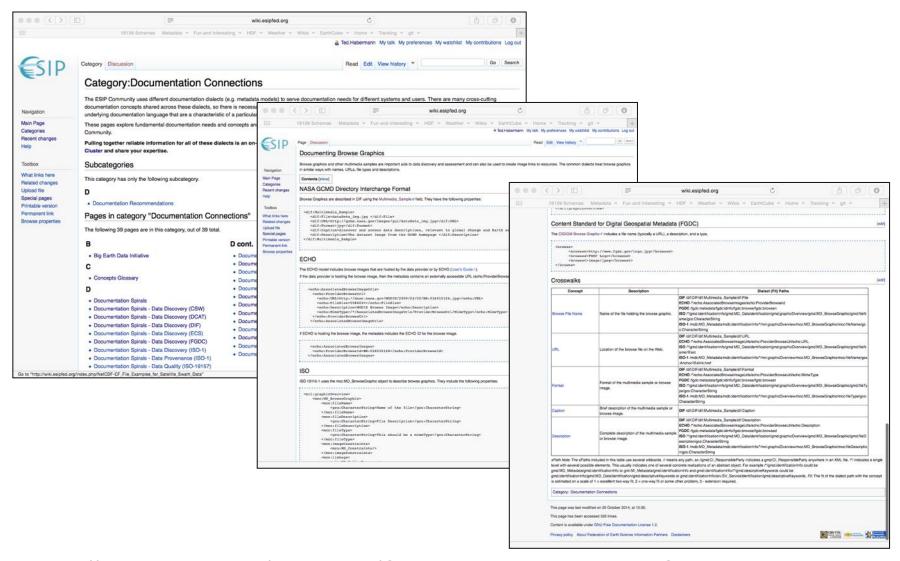
#### Signature Score Groups



#### Concept Guidance Links



#### **Guidance Documentation**



http://wiki.esipfed.org/index.php/Category:Documentation\_Connections

#### **Future Directions**

- How do I identify which records are missing concepts?
- What concepts are missing in a single record?
- Signature Score Sprints

DM\_PPT\_NP\_v02

13

#### How do I identify which records are missing concepts?

Concept	Concept Dialect Definition Link	Record Name							
		oli_aeri.xml, oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml,							
	Resource Format	oli_csphot.xml, oli_dl.xml, oli_dlprof.xmloli_ecmwfdiag.xml, oli_ecor.xml,							
Resource Format		oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml							
		oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml, oli_csphot.xml,							
	Coordinate Reference System (CRS)	oli_dl.xml, oli_dlprof.xml, oli_ecmwfdiag.xmloli_ecor.xml, oli_ghg.xml,							
Coordinate Reference System (CRS)		oli_gndrad.xml, oli_hsrl.xml, oli_htdma.xml							
		oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml, oli_htdma.xml, oli_irt.xml,							
	Association	oli_kasacr.xml, oli_kazr.xml, oli_masc.xml, oli_maws.xmloli_met.xml,							
Association		oli_mfr.xml, oli_mfrsr.xml, oli_mpl.xml, oli_mplcmask.xml							
		oli_csapr.xml, oli_csphot.xml, oli_dl.xml, oli_dlprof.xml, oli_ecmwfdiag.xml,							
	Author / Originator	oli_ecor.xml, oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xmloli_htdma.xml, oli_irt.xml,							
Author / Originator		oli_kasacr.xml, oli_kazr.xml, oli_masc.xml							
		oli_aeri.xml, oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml,							
	Publisher	oli_csphot.xml, oli_dl.xml, oli_dlprof.xmloli_ecmwfdiag.xml, oli_ecor.xml,							
Publisher		oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml							
		oli_aeri.xml, oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml,							
	Contributor Name	oli_cs <u>phot xml_oli_dl xml_oli_dlprof xmloli_ec</u> mwfdiag.xml, oli_ecor.xml,							
Contributor Name		oli ar Links to xPaths in							
	Sauras Citatian	on_ar							
Source Citation	Source Citation	oli gr particular dialect i_ecor.xml, oli_ghg.xml,							
Source Citation		oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml, oli_htdma.xml, oli_irt.xml,							
	Related Resource Citation	oli_kasacr.xml, oli_kazr.xml, oli_masc.xml, oli_maws.xmloli_met.xml,							
Related Resource Citation	Related Resource Citation	oli mfr.xml, oli mfrsr.xml, oli mpl.xml, oli mplcmask.xml							
Related Resource Citation		oli_csapr.xml, oli_csphot.xml, oli_dl.xml, oli_dlprof.xml, oli_ecmwfdiag.xml,							
	Resource Creation/Revision Date	oli_ecor.xml, oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xmloli_htdma.xml, oli_irt.xml,							
Resource Creation/Revision Date	nesource ereation/ nevision bate	oli_kasacr.xml, oli_kazr.xml, oli_masc.xml							
nessures or eation, nevision bute		oli_aeri.xml, oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml,							
	Publication Date	oli_csphot.xml, oli_dl.xml, oli_dlprof.xmloli_ecmwfdiag.xml, oli_ecor.xml,							
Publication Date		oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml							
		oli_aeri.xml, oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml,							
	Security Constraints	oli_csphot.xml, oli_dl.xml, oli_dlprof.xmloli_ecmwfdiag.xml, oli_ecor.xml,							
Security Constraints		oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml							
		oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml, oli_csphot.xml,							
	Parent Identifier	oli_dl.xml, oli_dlprof.xml, oli_ecmwfdiag.xmloli_ecor.xml, oli_ghg.xml,							
Parent Identifier		oli_gndrad.xml, oli_hsrl.xml, oli_htdma.xml							

## What concepts are missing in a single record?

Record Name	Concept Links to Dialect Definitions
oli_aeri.xml	Resource Format, Coordinate Reference System (CRS), Association
oli_amc.xml	Coordinate Reference System (CRS), Association, Author / Originator
oli_aod.xml	Association, Author / Originator, Publisher
oli_aos.xml	Author / Originator, Publisher, Contributor Name
oli_ceil.xml	Publisher, Contributor Name, Source Citation
oli_csapr.xml	Contributor Name, Source Citation, Related Resource Citation
oli_csphot.xml	Source Citation, Related Resource Citation, Resource Creation/Revision Date
oli_dl.xml	Related Resource Citation, Resource Creation/Revision Date, Publication Date
oli_dlprof.xml	Resource Creation/Revision Date, Publication Date, Security Constraints
oli_ecmwfdiag.xml	Publication Date, Security Constraints, Parent Identifier
oli_ecor.xml	Security Constraints, Parent Identifier,
oli_ghg.xml	Parent Identifier, ,

### Signature Score Sprints

Signature Score Group	Record Count	Record Name	Concept Links to Dialect Definitions
003	54	oli_aeri.xml, oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml, oli_csphot.xml, oli_dl.xml, oli_dlprof.xmloli_ecmwfdiag.xml, oli_ecor.xml, oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml	Resource Format, Coordinate Reference System (CRS), Association
0 0 4	32	oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml, oli_csphot.xml, oli_dl.xml, oli_dlprof.xml, oli_ecmwfdiag.xmloli_ecor.xml, oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml, oli_htdma.xml	Coordinate Reference System (CRS), Association, Author / Originator
0 4 4	16	oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml, oli_htdma.xml, oli_irt.xml, oli_kasacr.xml, oli_kazr.xml, oli_masc.xml, oli_maws.xmloli_met.xml, oli_mfr.xml, oli_mfr.xml, oli_mpl.xml, oli_mplcmask.xml	Association, Author / Originator, Publisher
051	85	oli_csapr.xml, oli_csphot.xml, oli_dl.xml, oli_dlprof.xml, oli_ecmwfdiag.xml, oli_ecor.xml, oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xmloli_htdma.xml, oli_irt.xml, oli_kasacr.xml, oli_kazr.xml, oli_masc.xml	Author / Originator, Publisher, Contributor Name
151	12	oli_aeri.xml, oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml, oli_csphot.xml, oli_dl.xml, oli_dlprof.xmloli_ecmwfdiag.xml, oli_ecor.xml, oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml	Publisher, Contributor Name, Source Citation
187	5	oli_aeri.xml, oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml, oli_csphot.xml, oli_dl.xml, oli_dlprof.xmloli_ecmwfdiag.xml, oli_ecor.xml, oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml	Contributor Name, Source Citation, Related Resource Citation
220	7	oli_amc.xml, oli_aod.xml, oli_aos.xml, oli_ceil.xml, oli_csapr.xml, oli_csphot.xml, oli_dl.xml, oli_dlprof.xml, oli_ecmwfdiag.xmloli_ecor.xml, oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml, oli_htdma.xml	Source Citation, Related Resource Citation, Resource Creation/Revision Date
300	2	oli_ghg.xml, oli_gndrad.xml, oli_hsrl.xml, oli_htdma.xml, oli_irt.xml, oli_kasacr.xml, oli_kazr.xml, oli_masc.xml, oli_maws.xmloli_met.xml, oli_mfr.xml, oli_mfr.xml, oli_mplcmask.xml	Related Resource Citation, Resource Creation/Revision Date, Publication Date

#### Acknowledgements



This work was partially supported by contract number NNG15HZ39C from NASA.



Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NASA or The HDF Group.

